Comparison of classification probabilities (based on EU classification cut points; i.e., 25, 200, 2000)

		Correct				More Stringent				Less Stringent			
LD50	slope	FDP	ATC	UDP	401	FDP	ATC	UDP	401	FDP	ATC	UDP	401
1.5	8.33	100	100	100	100	-	-	-	-	0	0	0	0
	2.0	100	100	100	99.9	-	-	-	-	0	0	0	0.1
	0.8	100	99.5	100	96.8	-	-	-	-	0	0.5	0	3.2
	0.5	100	96.6	100	95.1	-	-	-	-	0	3.4	0	4.9
50	8.33	99.9	100	100	100	0	0	0	0	0.1	0	0	0
	2.0	79.4	66.6	98.3	87.0	20.5	33.3	1.7	9.3	0.1	0.1	0	3.7
	0.8	9.2	39.3	92.1	67.0	90.7	56.7	7.9	21.9	0.1	3.9	0	11.1
	0.5	2.5	31.7	92.7	62.9	97.4	60.4	6.4	24.4	0.1	7.8	0.9	6.7
1500	8.33	0	99.6	98.5	97.9	0	0	0	0	100	0.4	1.5	2.1
	2.0	86.6	87.6	82.4	64.7	1.5	0.9	0	4.4	11.9	11.5	17.6	30.9
	0.8	24.2	58.6	75.3	48.8	75.2	31.0	0	6.9	0.7	10.7	24.7	44.3
	0.5	5.7	39.6	75.8	46.3	94.0	50.9	0	7.2	0.3	9.5	24.2	46.5
3000	8.33	100	97.1	99.9	99.9	0	2.9	0.1	0.1	-	-	-	-
	2.0	50.2	48.3	89.8	83.4	49.8	51.7	10.2	16.6	-	-	-	-
	0.8	2.5	22.3	85.2	73.5	97.5	77.5	14.8	26.5	-	-	-	-
	0.5	0.8	15.1	83.8	71.9	99.2	84.9	16.2	28.1	-	-	-	-

FDP and ATC are averaged across starting doses; FDP is the R=5 results; UDP is the LD50 results.

From the comparison table

For the most toxic substances (LD50=1.5), all seem to do well for various slopes.

For the substances with LD50=50, UDP does better than FDP & ATC as slope decreases (variance increases).

For less toxic substances (LD50=1500), UDP is still more often correct, but is more likely to underclassify as the slope decreases. (This may be a consequence of a poor (default) dose progression and an assumed (small) sigma.)

For the least toxic substances (LD50=3000), none underclassify, but the percentage overclassified increases dramatically with decreased slope.

Who did the work?

The analyses represent the work of:

401: Gregory Carr, USA
Proctor and Gamble

FDP(420): Nigel Stallard and Anne Whitehead, UK
University of Reading

ATC(423): Wolfgang Diener, Germany
BGVV

UDP(425): Elizabeth Margosches and Timothy Barry, USA
EPA

How was the work

All agreed to examine the behavior of the methods for substances with specific LD50/variance combinations. In order to have a common ground, all treated the data as lognormal, amenable to probit manipulations, and used the terminology LD50 and slope to designate the data characteristics. The EU classification cut offs (25, 200, 2000 mg/kg) were used.

The selection of doses is predetermined for FDP and ATC, but each proceeds differently according to start dose. *Calculated* probabilities of classification were provided for each start dose for the ATC and the FDP.

The selection of doses is arbitrary for UDP and 401 (in practice, informed by auxiliary information); 401 proceeds in a predetermined fashion once started; UDP proceeds differently according to each outcome. *Simulated* distributions of experimental LD50's were provided for three starting locations for the UDP and for three sets of dose arrays for the 401. From these distributions, probabilities of classification were *observed.*

All the analyses used LD50= 1.5, 50, 1500, 3000 and slope= 8.33, 2.0, 0.8, 0.5.

FDP analyses assumed 10 animals available at each dose tested. 401 analyses assumed 5 animals at each dose tested. ATC analyses assumed 3 animals at each dose tested. UDP used 1 animal at each dosing, but each dose may be visited repeatedly.

The summary table of comparisons was prepared by:

- Averaging FDP and ATC across starting dose.

  Successful classification by both the FDP and ATC becomes more dependent on starting dose as the LD50 increases closer to the greatest EU classification boundary (i.e., 2000) and the slope decreases.

  For LD50=3000, their classification at higher slopes is more dependent on starting dose, since the LD50 is greater than the boundary for the least stringent classification.

- Selecting the LD50 start for UDP.

  While probabilities of classification have not been calculated for the other starting doses, the spread of values in Table 3 of percentiles of the estimated LD50 indicates higher starting doses with decreasing slope give increased overestimation of LD50; lower starting doses with decreasing slope give increased underestimation.

  This is true for 401 as well, where the dose array bracketing the LD50 is the one in the summary comparison table.

- Using the FDP results for R=5

(R defines the proportionality of the evident toxicity curve).

While the probability of correct or more stringent classification is not much affected by this choice for the workshop analyses, the numbers of animals used are very different from those for R=50.

How could the alternative assays be improved?

•All will be improved by a sighting study, since all are affected by starting dose.

•To accommodate the harmonized classification system, the ATC and FDP will need changed prespecified doses.

UDP:
This method depends on the dose progression, which is related to the spread of responses, the length of the run, and the numbers of animals run per dose. Optimal dose progression has intervals equal to 1/slope; without information on slope, larger intervals increasing and smaller decreasing may provide better information. Multiple simultaneous starts (e.g., 3 trials concurrently) may provide better data. Two-parameter estimation is NOT necessarily better, since the estimate of sigma is still bound to be unreliable, and for the most part the LD50 estimate is similar.

FDP:
This method depends on the criterion for evident toxicity (which corresponds to the choice of R), the number of animals, and the prespecified doses at which it's performed. Whitehead and Curnow have noted a change in the last alone could give better concordance with LD50 results. Additionally, changing the number of animals responding to identify "less than 100% survival" or the number of animals tested for the base, can improve the performance.

ATC:
This method depends on the prespecified doses at which it's performed. These should conform with the desired classification system to give best performance.